US
PH PCT
010076W0
MAT DOSSIER

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
H04H 9/00, H04N 7/16

A1

(11) International Publication Number: WO 00/07310

(43) International Publication Date: 10 February 2000 (10.02.00)

(21) International Application Number: PCT/US99/16799

(22) International Filing Date: 22 July 1999 (22.07.99)

(30) Priority Data:
09/134,667 30 July 1998 (30.07.98) US

(71) Applicant: CBS CORPORATION [US/US]; 11 Stanwix Street, Pittsburgh, PA 15222–1384 (US).

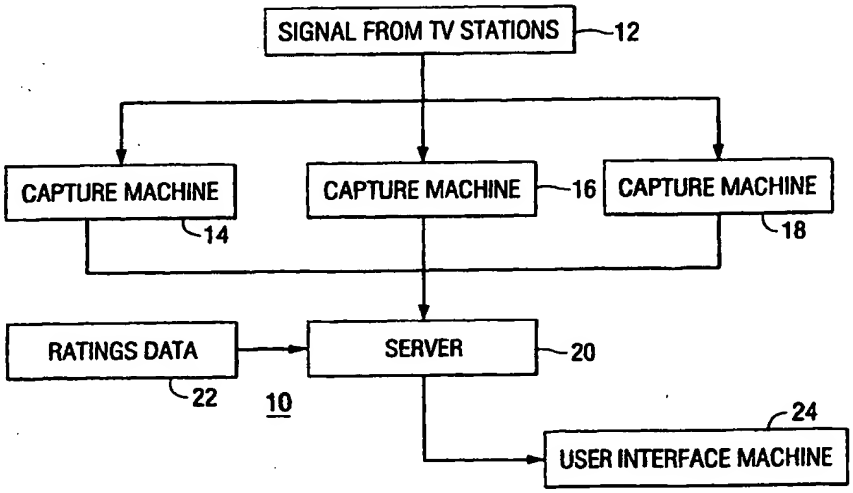
(72) Inventors: HULLINGER, Rick, A.; 5529 5th Avenue, Pittsburgh, PA 15232 (US). TROSKY, William, J.; 3001 Wedgewood Court, Murrysville, PA 15668 (US). MANDEL, Alan, F.; 263 Main Entrance Drive, Pittsburgh, PA 15228 (US).

(74) Agent: LENART, Robert, P.; Eckert Seamans Cherin & Mellott, LLC, 44th floor, 600 Grant Street, Pittsburgh, PA 15219 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: SYSTEM FOR ANALYZING TELEVISION PROGRAMS



(57) Abstract

A system that automatically captures one or more local news program broadcasts and separates the broadcasts into the individual news stories or segments. The system then compares the stories to historical data concerning the competitive characteristics of the stories for each station and determines the topic (local, national, crime, etc.), talent (newscaster 1, newscaster 2, etc.) and production (live, studio, voice–over–tape, etc.) characteristics of the stories. Other characteristics that affect the popularity and therefore the competitive characteristics of the broadcasts can also be displayed, such as pacing, average story length, news–to–advertisement ratio, broadcast ordering (news then weather then news then sports, etc.), etc. The characteristics are displayed in a visual format, such as a graph, with other historical data, such as show ratings that can be divided into increments such as 1/4 hour, and optionally with the actual video/audio broadcast allowing assessment of competitors local news broadcasts.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

SYSTEM FOR ANALYZING TELEVISION PROGRAMS

BACKGROUND OF THE INVENTION

Field of the Invention:

The present invention is directed to a system for analyzing television programs particularly local news programs and, more particularly, is directed to a system that captures a local news program broadcast, separates the broadcast into the individual news stories, determines the competitive characteristics of the stories which can include the topic, talent and production characteristics of the stories, combines the analysis results with historical data, such as show ratings, and provides the combined analysis along with the broadcast to a user allowing assessment of competitors local news broadcasts.

Description of the Related Art:

In today's competitive television environment where local television stations can derive a significant portion of their income from the sale of advertisements during half-hour news programs that are broadcast throughout the day, a small ratings increase can translate into the ability to significantly raise advertising rates for advertisements during these local news broadcasts. In the past, station managers and news program directors have had to rely on broadcast ratings (such as Nielsen ratings) and their own subjective experience in viewing competitors broadcasts to evaluate what aspects of the broadcasts contribute to improved ratings. What is needed is an objective analysis tool that quantifies the competitive characteristics about a broadcast allowing news directors to analytically determine what contributes to improved ratings.

Further, in today's environment keeping the attention of viewers is difficult. As a result, the factors that contribute to improved ratings will change over

time. What is needed is a system that will allow rapid analysis of a competitors broadcasts.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a system that objectively determines the competitive characteristics of news broadcasts.

It is a further object of the present invention to provide a station manager with improved competitive intelligence.

It is another object of the present invention to provide a news director with information about competing news broadcasts that has not previously been available.

It is also an object of the present invention to provide a system that helps a news director to optimize newscast ratings with respect to competitors.

It is a further object of the invention to provide a system that allows determination of the topic, talent and production characteristics of a broadcast.

It is an object of the present invention to provide a system that allows determination of how the topic, talent and production affect ratings.

It is another object of the present invention to digitize television broadcasts to allow display of the video on a computer with the competitive analysis data.

It is another object of the present invention to provide the ability to compare or correlate story content and/or competitive characteristics with ratings data.

It is a further object of the present invention to provide a system that allows monitoring of competition on a relatively real time basis.

It is a further object of the present invention to provide instant access to the video and audio content of specific news stories.

The above objects can be attained by a system that captures a local news program broadcast, separates the broadcast into the individual news stories, and determines the topic, talent and production characteristics of the stories by comparing the text of the broadcast with statistical information accumulated about prior broadcasts. The characteristics are combined with other historical data, such as show ratings. The system provides the combined analysis along with the broadcast

itself to a user through a graphical user interface allowing an objective assessment of competitors local news broadcasts.

These together with other objects and advantages that will be subsequently apparent, reside in the details of construction and operation as more fully hereinafter described and claimed, reference being had to the accompanying drawings forming a part hereof, wherein like numerals refer to like parts throughout.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts the hardware of the present invention.

Figure 2 illustrates data flow of the capture process within each capture machine of figure 2.

Figure 3 is a block diagram of the operations performed in the present invention.

Figure 4 depicts the operations of the capture process 52.

Figure 5 depicts the operations of the preparser process 54.

Figure 6 is an example of a score tree.

Figure 7 depicts the operations of the parser process 58.

Figure 8 shows the flow of a manual classification process.

Figure 9 is a flow diagram of a statistical process that determines values in score tables.

Figure 10 depicts a preferred structure of the database of the present invention.

Figure 11 illustrates types of classification charts.

Figure 12 illustrates charts with other types of interfaces being simultaneously displayed.

Figure 13 illustrates a ratings chart.

Figure 14 depicts a main user interface screen.

Figure 15 illustrates a view of information about a particular television station.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is directed to a system that analyzes television program broadcasts, particularly newscasts, and determines the correlations between

content of the news (what was presented, who presented it, and how it was presented) and the ratings for the broadcast. This information, along with the broadcasts themselves, can then be used to optimize future newscasts to increase ratings. A television program broadcast can include one or more television programs and can be transmitted over a broad variety of media including a traditional airwaves broadcast, a cable broadcast, and a digital broadcast over a network, such as the Internet or any other medium suitable for distributing television type programs. A broadcast can also be initiated by the broadcaster or by the consumer when a computer user, through a web browser, requests the download of a television program. A television program can include newscasts and other types of television programs as well as other types of video/audio material suitable for viewing by a user.

The system 10, as depicted in figure 1, receives one or more television signals 12 from one or more local television stations. The reception can be via air waves, cable, digital network, magnetic media and any other media suitable for inputting video/audio material into the system 10. In the Pittsburgh market the local stations are KDKA, WTAE and WPXI. The signals are routed to one or more capture machines 14, 16 and 18, preferably, with a capture machine corresponding to each of the signals to be processed. However, it is possible to have a single machine performing the task of several capture machines although such is not preferred. These capture machines 14, 16 and 18 capture the video and audio of the broadcast as well as the closed-caption (CC) text broadcast with the program. The machines 14, 16 and 18 break the news broadcast into stories (or segments) and classify each story by the dominant topic, such as weather or sports, by talent, such as the person presenting the story, and by production, such as studio or live. Stories generally have the same characteristics (topic, talent and production) and where one of these characteristics changes the story generally changes. For example, a first video segment may be a local segment, by newscaster #1, and live. If the second segment is also a local segment, by newscaster #1, but the production type is studio, then a new story has started with segment 2. However, if the topic, talent and production types are the same for both segments, then they will be considered parts of the same story.

RNSDOCID: <WO 0007310A1 I >

Each of the capture machines 14, 16 and 18 preferably is based on an IBM compatible personal computer (PC). The system preferably uses the Windows 95/98 operating system. The capture machines 14, 16 and 18 could alternatively include the hardware necessary to capture the video and compress it into an MPEG-1 video stream.

The broadcast data generated from the analysis is transferred to a server 20 which also receives ratings and share data 22 from a conventional source such as Nielsen overnight reports. The server 20 is preferably an IBM Compatible PC. The system 20 preferably uses the Windows NT operating system.

The broadcast and segment data, the video and audio data, and the ratings data stored on the server 20 are made available to a news program manager through one or more user interface machines 24. The user can play the video and audio, and review the data in the form of charts, etc. Each user interface machine 24 is preferably an IBM Compatible PC. The system 24 preferably uses the Windows NT operating system.

If processing needs to be faster, a special processing machine (not shown) can be positioned between the capture machines 14, 16 and 18 and the server 20. This processing machine would perform the analysis discussed herein while the capture machines 14, 1G and 18 would be dedicated to the capture process. This machine could be based on Digital Equipment Corporation (DEC) AlphaStation 500 MHz Processor. This additional system preferably would use the DEC UNIX operating system. Of course, as previously mentioned the entire system could be implemented in a single system 10 based on a midsize computer such as an ALPHA machine made by Digital Equipment Corporation.

Figure 2 illustrates the signal and data flow that occurs within each of the capture machines 14, 16 and 18. The signals 12 from the television broadcasts are supplied to the closed-caption capture board or unit 32 previously mentioned. This board 32 outputs a closed-captioned text file 34, the use of which will be discussed in more detail later, and separately outputs the video signal to the video capture board or unit 36 previously mentioned and the audio signal to the sound board or unit 38 previously mentioned. These two boards together produce a video file 40.

As depicted in figure 3, the invention includes several stages of processing. The first stage is capture processing 52 which results in the files 34 and 40 previously mentioned. The capture process 52, which will be discussed in more detail later, essentially starts the capture of the video signal at the proper time, creates the appropriate files 34 and 40, and then starts the preparser 54. The closed - captioned raw data file 34 is supplied to the preparser process 54, which will be discussed in more detail later, essentially removes meaningless characters from the closed-captioned text and adds information such as broadcast slot and related file identifiers. This creates a processed closed captioned text file 56 which is supplied to a parser process 58 and stored in the server 20. The parser process 56, which is typically executed immediately after the preparsing process 54 and which will be discussed in more detail later, essentially reviews the closed-caption text 56, divides the broadcast into stories or segments, and, determines the topic, talent and production values for each segment. This information along with the closed-caption (CC) text is stored in database tables and associated files 60, 62 and 64 in the server 20. The user at the user interface machine 24 can access and display the data stored in the server 20 (classification, ratings, pacing, average story length, etc.) along with playing the video/audio to see what components of the broadcast affect the ratings for a broadcast.

The capture process 52, as depicted in figure 4, once the hardware boards 32, 36 and 38 have been initialized, waits 72 for the start time which coincides with the beginning of a target news broadcast. Once the start time has been reached, the status of the system is updated 74 to "capturing" status, so that a user will be informed that capture has started, if an inquiry is made. Next, the text capture is started 76 by board 32 followed by the start 78 of the video/audio capture by boards 36 and 38 at a desired capture video frame rate and resolution, such as 5 frames/second at 160x120. When the end of the broadcast time has been reached, the text and video/audio capture is stopped 82/84 and the status is updated 86 to "idle." The files, with appropriate information concerning the channel, time of capture, etc., are then transferred 88 and the pre-parsing process is then started 90.

The preparser process 54, as depicted in figure 5, starts with the raw closed-caption data file 34 and essentially converts human friendly data into machine

friendly data. The header of the file 34 is used to obtain 112 the date and time of capture as well as the station number. The last record in the file 34 is read to obtain the ending time of the text capture. Next, the broadcast slot ID of the broadcast is determined by accessing the (relational) database which shows the links between broadcast start and end times and broadcast slot ID's. A new file 56 (the processed or preprocessed data file) is then created 118 and the header is updated with the ID as well as the date and frame rate of the capture. The offset in frames from the start of the video capture to the start of the text capture is determined 120. This allows the text to be correlated to the exact frame(s) in which it is produced. The offset is determined by obtaining a difference (which can be negative) between the time of the first line of text data and the video capture start time. This difference is multiplied by the frame rate. This value is stored in the header of the new file 56. Next, a line of text is read 122 and the line of the text is truncated 124 if it is longer that a predetermined length, such as 80 characters. Each of the characters in a text line is also examined to determine if it is a valid character (not valid unless the ASCII value is 32122) and if not the character is replaced 126 with a space. Then the text is written 128 to the output file 56 with the appropriate time stamps. A determination 130 is then made as to whether the end of the data has been reached. If not, the process continues, and if so, the process is finished 131.

The parser process 58 uses two types of tables: a vocabulary table and a score table. The contents and structure of these tables will be discussed before the parser process 58 is discussed. How these tables are created will be discussed later herein.

There are essentially at least three vocabulary tables, one for each length of phrase that will be examined during the parsing. That is, a one-word vocabulary table holds phrases of one word (see the example below), a two-word vocabulary table holds phrases of two words and a three-word vocabulary table holds phrases of three words. The phrases are stored in alphabetical order for fast searching, and each phrase entry contains three pieces of data: 1) The text of the phrase; 2) a unique phrase ID, which is used to look up the phrase in the designated score tables; and 3) an array of score table identifiers (IDs) indicating which score tables/files contain the phrase. The table IDs are references to actual files (names),

so when the parsing process 58 is looking up the score data for a phrase, the process 58 knows which files contain the phrase, and can limit a search to the designated files.

Example One-Word Vocabulary Table:

Phrase	Phrase ID	Score1 TableId	Score2 TableId	Score3 TableId
Crime	28	1	2	5
Crimes	29	2	3	12
Criminal	30	1	2	8

In the above example, the word "crimes" has a Phrase ID of 29 and can be found in score tables 2, 3 and 12.

There are N score tables where N is the number nodes in the scoring table tree. The tree will be discussed in more detail with respect to figure 6. Each table (see the example below) contains a list of phrases, their occurrence frequency and scoring data. The phrases are stored in ascending numerical order by PhraseID for fast searching. Each entry contains the following pieces of data: 1) The PhraseID that refers back to the actual text in a vocabulary table. 2) The total count which is the total number of times that this particular phrase has been seen in closed-caption (CC) text before. 3) The topic counts, which are then the number of times that this phrase has been seen in the CC text that was about the given topic (Topic1, Topic2, ...). 4) The talent counts which are the number of times that this phrase has been seen in CC Text that was presented by the given talent (Talent1, Talent2, ...). 5) The production counts which are the number of times that this phrase has been seen in CC Text that was presented with the given production type (Production1, Production2, ...) A typical set of topics could include: local, national, international, sports, weather, advertisement, tease, other and unknown. A typical set of production values could include: live, tape, studio, other and unknown. The set of talent (that is, the people involved in the broadcasts) is dependant on the market being monitored, and will even change within that market as reporters start, quit and change stations.

Table I

Phrase ID	28	29	30	31
Total Count	194	17	241	100
Topic 1	81	5	7	18
Topic 2	23	8	203	41
Topic N	17	0	28	22
Talent 1	44	15	77	19
Talent 2	21	2	23	51
Talent N	19	0	14	2
Production 1	130	12	120	0
Production 2	52	4	7938	
Production N	5	0	6	12

In Table I, for phrase 29 ("Crimes") the total count is 17, the topics 1, 2 ... n have scores of 5, 8 and 0 respectively, talents 1, 2 ... n have scores 15, 2 and 0 respectively and production values 1, 2 ... n have scores of 12, 4 and 0 respectively.

A constraint, which follows from the way in which the score table is generated (to be discussed later herein), is preferably placed on the entries in the score table. The sum of the topic counts equals the sum of the talent counts, which equals the sum of the production counts, which equals the total count. That is to say that every time a phrase is seen in CC text, it is recorded in the appropriate score table(s), the total count is incremented, and one and only one of the topic counts, talent counts and production counts is incremented.

As a more concrete example, if a particular news program manager or news director was only interested in recording four topics (Local, National, Weather and other), only had 3 Talents (Sam D., Dan R. and P. Jennings), and only cared about three production types (Live, Studio and other), one of the score tables might look something like below.

Table II

Phrase ID	28	29	30
Total Count	100	9	25

Local	70	9	1
National	25	0	12
Weather	2	0	12
Other	3	0	0
Sam D.	35	5	10
Dan R.	35	1	8
Peter J.	30	3	7
Live	28	8	5
Studio	72	0	9
Other	0	1	11

As previously mentioned there are N score tables. The number N can be determined using a score table tree such as illustrated in figure 6 which depicts ten score tables. Each node of the score table tree represents a score table/file that contains scoring data derived from a specific subset of all of the closed-caption text previously processed by the system. When the present invention is implemented at a new location, the tree would be typically set up with three levels. The top-level node 132 is the general score file/table and holds scores for every phrase encountered in all the news stories included in the system. The second level nodes 134, 136 and 138 hold the station score files/tables, one file for each station that the system is recording. For example, the score table for node 134 holds all the phrases for every news story run by station KDKA included in the system. The third level nodes 140, 142, 144, 146, 148 and 150 hold scores for all phrases the system has recorded. The tree of figure 6 is just an example and new or additional nodes with more specific definitions could be added to the tree below the current leaf nodes at any time. For example, it is possible to add another level to the tree below the broadcast slot level that contained two nodes below each broadcast slot node: one with scores for phrases heard in the first 15 minutes of the broadcast, and the other with phrases heard in the last 15 minutes of the broadcast.

The parser process 58, as depicted in figure 7, starts by creating 162 a text segment list where each segment corresponds to a line of text in the closed-caption (CC) text file 56. With each line of text being an individual segment

the process 58 essentially scores each segment and examines each segment to see if it can be combined with the segment directly in front of or behind it using the scores. If the segment can be combined because they have the same topic, talent and production, then they are combined. Once the segments are combined they are rescored because the score of the new segment can change due to the combination.

In determining whether segments should be combined or maintained as separate stories, the first step is to score 164 all the segments. This is accomplished by scoring the three-word phrases, the two word phrases and the one word phrases in the segment. Of course it is possible to limit the system to only one or two word phrases or to additionally allow four, five or more word phrases if desired. To do this the phrase is looked-up in the vocabulary table to obtain the phraseID and the score tables pointers for the phrase. The most detailed score tables are accessed based on the lowest table of the tree found in the score table list. For example, if the phrase appears in a broadcast slot table (say table 150 of figure 6) it is accessed, next a station table (such as table 138 in figure 6) is examined for the phrase and if not found the general score table (say 132 of figure 6) is used. The segment score for each of the topic, talent and production values found in the table is updated (accumulated).

Once two segments are scored they can be combined 166. If the topic, talent and production values for two segments match, they are combined into a single segment. The system also combines segments by looking at tease type segments between similar segments. If a segment is a tease segment it is combined with the segment in front of it or the segment behind it based on the similarity of the topic, talent and production values of the tease and the adjacent segments. The current segment is combined with the most similar segment. To separate segments 168 the system essentially looks for time gaps (a minimum of 30 seconds) in segments that are at least a minimum length in time (preferably 5 minutes) and do not have a segment topic classification of other. The combined and separated segments result in a revision of the segments list.

Once the segments are combined or separated, they are rescored 170. When phrases are scored, the system attempts to score the longest phrase possible under the assumption that longer phrases are more specific, and will provide more

accurate scoring information. If, when scoring, the system scores all of the words and phrases in a segment up to, but not including the final word, then the system can only look up the scores for the final one-word phrase, because there's only one word left in the segment and scoring between segments is not performed.

When a segment is combined with the segment following it, that single word is no longer at the end of the segment, but somewhere in the middle, and it may now be part of a two or three word phrase. If this is the case, then two things change. 1) The new two or three word phrase must now be accounted for in the scoring, because it may have drastically different scores than the one-word phrase did. 2) Everything following the new phrase must be re-scored, because if the new phrase uses the first one or two words of the second segment, then the phrase matching for the rest of the segment will turn up different phrases and score differently. As an example, assume that the system is presented with the sentence "A Pittsburgher walked to the Statue of Liberty to visit the President of France" which appears as two adjacent segments in the CC text. The two segments being: segmentl = "A Pittsburgher walked to the Statue"; and segment 2 "of Liberty to visit the President of France.'' Assume also that the phrases in the segments are scored as follows:

Phrase	Local	National	International
A Pittsburgher	0.7	0.1	0.1
walked to the	0.3	0.0	0.0
Statue	0.2	0.2	0.2
of	0.0	0.0	0.0
Liberty to visit	0.7	0.2	0.1
the President of	0.1	0.7	0.2
France	0.1	0.3	0.6
Statue of Liberty	0.0	0.7	0.3
to visit the	0.0	0.0	0.0
President of France	0.0	0.0	1.0

With this phrase scoring segment 1 has a score of Local = 1.2, National = 0.3 and International = 0.3 while segment 2 has a score of Local = 0.9, National = 1.2 and International = 0.9. If segment 1 and segment 2 are

combined simply by adding their scores and not re-scoring the new segment, the scores would be: Local = 2.1, National = 1.5 and International 1.2. Because the local topic has the highest score the new segment would be given a classification of "Local". However, if the segment is re-scored, we have:

New segment:

"A Pittsburgher walked to the Statue of Liberty to visit the President of France"

New phrases:

A Pittsburgher, walked to the, Statue of Liberty, to visit the, President of France

Score:

Local: 1.0, National: 0.8, International: 1.4

The rescoring of the newly combined segment would result in the segment being classified as "International". This same score changing may occur when the segments are broken apart, so segments must be re-scored every time they are combined or split up.

Once the segments are rescored, the segment list is updated and the stories and their classification data are stored in the segment table of the database. Also, the broadcast data is stored in the database's broadcasts table and the corresponding closed-captioned text is also stored in the CC text file 64.

Once the stories for one or more broadcasts have been classified and the ratings data for the broadcast has been updated, the user can access and display the data. However, before discussing the display of the data, a description of how the vocabulary and score tables are created will be provided.

To create the various tables, the closed captioned text of file 56 needs to be classified to create training data and this is performed by a manual classification process 190 illustrated in figure 8. This can be done immediately after the broadcast is captured, if a real time type analysis is desired, or at some later date. To perform the manual classification the system loads 192 the video/audio file 40 into a single video player, which will be discussed in more detail later herein, and loads the CC text file 56 into a CC text viewer. The video is played and viewed by an individual, called a classifier for convenience, who will perform a manual

classification of the text. The classifier determines the starting and ending points of a story or segment in the text by viewing and essentially marking 194 the text using a conventional blocking method much like the blocking method used to mark (or highlight) text to be copied/cut in a word processor operation. The classifier then enters 196 the appropriate classifications for the segment by indicating the topic, talent and production for the segment. The system then records 198 this text as a story segment along with the classification, and the starting and ending lines of the text. The classification is then recorded 200 for each segment and the segment and associated classification is stored 202. A segment holds all of the information about topic, talent and production. Each line of cc-text stored and sequentially numbered. The segment holds a topic, talent and production identifier, and a cctext start ID and cc-text end ID for the segment. Every line of cc-text between the start line and the end line are a part of that segment, and therefore have the same topic, talent and production values. Note that the classifier can be an automated, computer based classifier such as an expert system rather than a person.

Once the story has been classified (the training data created) the vocabulary and score tables can be created. It is preferable that the stories for some period of time, such as a week, be accumulated before the tables are created or updated. However, the statistical process can be run after each broadcast is classified when an up-to-the-minute database is desired. The statistical information process 210, as illustrated in figure 9, starts with parsing 212 the text into all of the separate words. Then, all of the 1, 2 and 3 word phrases are constructed 214. For example, the text "the door is now closed" becomes the 12 phrases: the; door; is; now; closed; the door; door is; is now; now closed; the door is; door is now; is now closed.

Each phrase is then looked up 216 in the vocabulary table. If the phrase is determined 217 not to be in the vocabulary table, a new entry is made in the vocabulary table, a phraseID is assigned, a pointer to the appropriate score table is created and the scores/counts (total and appropriate topic, talent and production) of the score table for the slot of the broadcast is updated from the information for the text entered by the classifier. Note that the phrases in the vocabulary table are in alphabetical order so sorting and other housekeeping operations for the vocabulary

table responsive to the new phrase may be required. Preferably, all of the new phrases would be accumulated in a new phrase file. At the end of the statistical process the files are combined/merged and sorted into the preferred alphabetical order.

When the phrase is in the vocabulary table, the score table pointers are used to address 218 the appropriate score table for the slot of the broadcast and the scores in the table are updated 220. If the phrase is not in the table, a new phrase is created 222 and stored, and then the scores for the new phrase are updated.

The database of the present invention preferably has a structure as depicted in figure 10. This data structure, which is typically stored on a computer readable medium and facilitates the efficient execution of the process described herein includes a pointer directed set of tables which include the vocabulary tables and score tables previously discussed.

As previously mentioned, once the stories for one or more broadcasts have been classified and the ratings data for the broadcast has been updated, the user can access the data through the user interface device by playing the broadcasts individually or simultaneously using a specialized video player. The user can also simultaneously or separately display the text of the broadcast and the ratings data in the form of charts. The data charts will be discussed first followed by the video players.

Several different types of charts are provided by the present invention. The charts themselves are created and displayed via conventional chart creation techniques, and the interaction of the charts with other objects, such as the video players, etc. is discussed in more detail herein.

Figure 11 illustrates the types of charts that are used to display the classification data for the stories of a particular broadcast slot while figure 12 shows charts with interfaces to other types of data. Figure 11 provides a view that allows the user to compare the average topic, talent and production for two or more stations. When the user moves the cursor over a particular bar or pie wedge, then the chart displays a bubble label giving the user more information about the particular data object. Left clicking the mouse when the cursor is over a data object

displays a context menu that allows the user to choose several ways to view the data of the object in more detail. Double clicking the mouse when the cursor is over a data object opens a video player so that the user can watch the individual stories that are represented by the data object.

Figure 13 illustrates a rating chart. The ratings view of figure 13 allows the user to look at ratings and/or share data for the time period specified in the Start Time and End Time fields in the toolbar. The chart of figure 13 shows ratings share data, along with a linear-fit of the data to aid in analyzing ratings trends. When the user right-clicks on the background of any chart, the chart properties dialog is displayed (options in the dialog change slightly with chart type (bar and pie or ratings plot)). The ratings chart which may include the ability to look at any combination of ratings and share data, actual data or linear fit approximations, as well as control over placement and visibility of the title and key. Other options include the ability to view the half-hour average ratings data, the change in viewership across a 15-minute break or the change in viewership from the lead-in program. Similarly, the user can use the chart properties dialog to access axis customization tools. Here the user can establish the range of the X and Y axes, change the spacing between tick-marks on the axes, and turn grid lines in the X and Y direction on or off.

All of the charts and graphs in the system of the present invention, regardless of their style (bar chart, pie chart, plot ...) or how they were created, are derived from a single chart object, so they all share the same set of attributes and interactions.

In the chart the term "data area" refers to any particular piece of data in the chart (a single bar, pie wedge, or plot point) and the term "background" refers to any part of the chart that is not a data area.

Any time the phrase "stories that make up the data" is used, it means that the application queries the database for a list of stories that fit the specified criteria. If the chart shows the topic breakdown for KDKA's evening news from 01/01/97 to 01/31/97 and the user double-clicks on the bar representing the national news, then "all the stores that make up that data" would be all of the stories

that ran on KDKA's evening news from 01/01/97 to 01/01/31 that were labeled as national news.

A single video player is used to view a video(s) from a single television station. The creation of the video player and the conventional functions such as play, rewind, etc. are implemented conventionally. The additional functions needed by the present invention are discussed in detail herein. A single video player is created when the user double-clicks on chart data, or selects "Video" from a chart's context menu. The video player is then loaded with the stories that are represented by the chart data object, and the user can watch each of the stories sequentially. Most of the video player's controls behave like conventional VCR and computer video controls, with the exception being the "track forward" and "track backward" buttons, usually only seen on CD player controls. These two buttons allow the user to hop forward (or backward) to the start of the next story. When using a single video player, the user is capable using of the following functions which preferably appear as buttons on a player window: Play, where the video is conventionally played; Fast-Forward, where the video is conventionally moved forward an accelerated rate; Rewind, where the video is conventionally rewound; Pause, where the video is conventionally paused; and Stop, where the video is conventionally stopped.

Volume on the audio portion of the video can be controlled. Mute is where the sound is turned off. Next-Segment and Previous-Segment access causes the player to instantly jumps to the start of the next or previous segment. Non-Linear Playback/scanning operates in two modes. 1) If the video player is loaded with stories that are not chronologically continuous, the player will play/fast-forward/rewind through the stories as if there were no gaps between them. 2) If the video player is loaded with stories that reside in separate files (akin to being on separate videotapes), the player loads and unloads files transparently and play/fast-forward/rewind as if the

A multi-video player (see figure 12) is created when the user selects the Newscast Video button and more than one station (Station Buttons) are currently selected for analysis. The multi-video player loads broadcasts from all of the selected stations, and then synchronizes each of them with respect to time. The user

can then manipulate each of the videos individually using the controls directly below each window. Control over which video is providing the audio feed is achieved with the speaker buttons directly to the right of each set of video controls. Clicking on a speaker icon gives the associated video player audio control. The lower set of "spanning" controls allow the user to control all of the video players together, playing, fast-forwarding and rewinding while keeping the videos all in sync. If the videos do get out of sync (either by manipulating a single video player or by using the story forward/back buttons on the spanning control, they can be re-synchronized using the sync button (the button displaying a clap-board to the left of the spanning controls). When the sync button is pressed, each of the videos is moved to the time of the video that has audio control. When using a multiple video player, the user is capable of using the following functions which preferably appear as buttons that are associated with the windows but separate from the windows for the individual players: 1) all functionality of the single video player; 2) ability to play and manipulate n videos simultaneously; 3) Play, Fast-Forward, Rewind, Pause and Stop all videos together with the push of a single button; 4) toggle audio feed between videos; and 5) synchronize instantly sets each video to the same date and time based on the video that currently has audio control.

Interactions listed for the single player work the same way for the multi-player. Using one of the "spanning controls" the system behaves as if the action was taken on each of the individual players simultaneously.

If the user wishes to view the actual text of a broadcast it can be viewed using the manual classifier. To aid the user in interacting with the processes of the present invention previously discussed, the invention includes additional displays as will be discussed below. The main screen is preferably the first screen that the user sees after logging into the system. It allows the user to select the stations, date range and newscasts that he or she wishes to analyze, and select the view that he or she wishes to see.

A Transcript button displays the closed-captioned text by opening a text window showing a.transcript of the broadcast, labeling each talent transition with the name of the talent who is beginning to speak. The three buttons shown in the main screen of figure 14, which look like calendars with one day, one week and

one month highlighted, allow the user to quickly change the date range to a single day, week or month, respectively. When one of these buttons is clicked, the Start Date is updated to be one day/week/month before the End Date.

The next four buttons are the Station Buttons. They show the channel numbers of stations in the local market which are being monitored. The user picks which station(s) he or she wishes to view data from by selecting one or more of these buttons. In the screen shot of figure 14, the user has chosen to view data from only channel 2.

The next three buttons are the Broadcast Buttons. Like the station buttons, they represent the broadcasts in the local market which are being monitored. The user picks which broadcasts he or she wishes to analyze by selecting one or more of these buttons. Each button shows a clock face showing a particular hour of time, where blue hands indicate p.m. and orange hands represent a.m. In the screen shot of figure 14, the broadcast buttons represent 6pm, 11pm and 12am, and the user has selected the 11 o'clock news to analyze. The "OK" button is used to refresh the currently active view when the date, broadcast or station buttons have been updated. Finally, the help button provides access to the help system.

The second row of tools give the user additional control over the stations, broadcasts and dates to analyze. The system of the present invention also provides several menus.

The present invention also provides the ability to control the dates upon which data is analyzed using an interface as depicted in figure 14. The date selection dialog allows the user to select the start and end dates for analysis. Clicking on a date in the top calendar (window) establishes a start date, and the lower window establishes the end date, and all of the highlighted dates will be used for analysis. If the user wishes to exclude particular days of the week, Day Configuration check boxes to the right of the calendars can be used to toggle specific days on or off. The user may also choose to toggle specific days on or off by left-clicking on the date, allowing complete flexibility when choosing dates to analyze.

A detail view, as illustrated in figure 15, is created when the user selects "Detail Info'' from a chart's context menu. This view provides the user with

a more detailed visual breakdown of the data represented by the data object the user clicked on.

A station detail view, as illustrated in figure 15, provides a view of all of the pertinent data for a single station. The topmost fields give information about the station, the broadcast(s) used to generate the data, the ratio of time spent in news to time spend in ads, and story pacing info. The middle of the view provides an overview of the ratings, and the bottom three charts provide the average topic, talent and production breakdown for the station. All of the charts in this view can be updated and customized in the same manor as the other views, and the chart's context menus can be used to obtain more information about specific data.

A segment list is created when the user selects "Segment List'' from a chart's context menu. This view provides a textual breakdown of the segments represented by the data object that the user clicked on. The stories shown in this view can be sorted by any of the columns by clicking on the column header.

Additionally, the user can double click on any of the stories, and a video player will open to show the story.

A Capture Time Selection Dialog allows the user to select which times and stations to capture. To select a capture time, the user highlights the time slot on the grid (which behaves like an MS Excel worksheet), and then selects the days to capture and gives the new capture slot a name. Once the information has been specified, the user can click on Add/Update, the new information will be sent to the database and the system will be ready to capture the new broadcasts. Similarly, by highlighting an already existing capture slot, the user can re-name, update or remove capture slots from the system.

The present invention has been described as having certain capabilities and features. The present invention can also include the following additional charting capabilities: 1. The ability to view ratings data and data about a specific topic, talent or production on the same chart - time spent in local news vs. ratings. 2. The ability to provide the facilities to set one chart as a benchmark and have the related charts show their values as deltas from the benchmark value. 3. The ability to use queries to generate charts - charts that show ratings for the days that the lead story was weather related, or charts that show the topic breakdown for

broadcasts that lost ratings at the 15 minute break. 4. The ability to generate formatted transcripts for any broadcast. 5. The ability to chart the story or content overlap between the any two (or more) newscasts (i.e. how much of the 11 o'clock news was the same as what we showed at 6 o'clock?). 6. The ability to show timeline charts depicting the specific times in a broadcast that a particular topic/talent/production was used.

The invention can include the following additional video player capabilities: a) skip forward / backward to the next story of a specific topic / talent / production - go to the next "live" story; b) the ability to synchronize multiple players to a particular segment - move all three players to their weather stories; c) the ability to load a multiple video player with segments of a particular topic, talent or production (currently multiple video players can only be loaded with entire broadcasts, not a subset of segments); and d) the ability to print any of the charts and views that the software creates.

The present invention has also been described with respect to topic, talent and production being broken down into rather broad categories such as topic being broken down into local, national, etc. It is desirable to break stories down into an even finer granularity such as local crime stories, local-fire stories, state-sports-baseball stores, etc. The invention has also been described with respect to using tables of 1, 2 and 3 word phrases. The phrases can also be 4 or more words of desired. To improve the performance and resource utilization characteristics of the system common words such as conjunctions, prepositions and articles can be eliminated when scoring and analyzing. Neglected phrases, that is phrases that have only been encountered once or twice in a year can also be removed to enhance the system. The invention has been described with respect to rescoring segments after they have been combined, that is, a single cycle of counting and rescoring. It is possible for this to be done a number of times. The techniques of the invention can be used to monitor other types of shows (network news, prime time). The invention can further be used to watch the broadcasts all day long and pick out and categorize all of the stations promotional spots ("coming up tonight at 11!"). The user would then be able to do analysis similar to that for the news. Also the invention can be

used for categorizing advertisements that run on the monitoring stations, and providing analysis of the ads.

The many features and advantages of the invention are apparent from the detailed specification and, thus, it is intended by the appended claims to cover all such features and advantages of the invention which fall within the true spirit and scope of the invention. Further, since numerous modifications and changes will readily occur to those skilled in the art, it is not desired to limit the invention to the exact construction and operation illustrated and described, and accordingly all suitable modifications and equivalents may be resorted to, falling within the scope of the invention.

What is claimed is:

1. An apparatus for analyzing television broadcasts, comprising:

a capture system capturing a designated television broadcast; and

an analysis system analyzing the broadcast and determining competitive characteristics of the broadcast.

2. An apparatus as recited in claim 1, wherein the competitive characteristics comprise one of broadcast topic, talent and production.

3. An apparatus as recited in claim 1 where said analysis system separates the broadcast into individual stories.

4. An apparatus as recited in claim 1, wherein said analysis system determines the competitive characteristics by comparing spoken text of the broadcast with statistical information accumulated about prior broadcasts.

5. An apparatus as recited in claim 1, further comprising a user interface displaying the competitive characteristics of the broadcast for a user.

6. An apparatus as recited in claim 5, wherein said interface graphically displays the characteristics.

7. An apparatus as recited in claim 5, wherein said interface simultaneously displays the characteristics and the broadcast.

8. An apparatus as recited in claim 5, wherein said interface simultaneously displays the characteristics and ratings data for the broadcast.

9. An apparatus as recited in claim 1, wherein said analysis system scores all segments of text, determines whether the segments can be combined and rescores the combined segments.

10. An apparatus as recited in claim 9, wherein said apparatus scores and rescores by comparing the text to a vocabulary table and determining a

score table from the vocabulary table and accumulating scores for the text from the score table with the highest score for each classification category becoming a classification for the text.

11. An apparatus as recited in claim 10, further comprising a score table tree of the statistical information.

12. An apparatus as recited in claim 1, wherein said capture system comprises a capture machine for each television station being analyzed.

13. An apparatus as recited in claim 12, wherein each capture machine comprises:

a closed-caption unit receiving capturing closed captioned text for the broadcast; and

a video/audio capture unit converting a video of the broadcast into digital images and an audio of the broadcast into digital sound.

14. An apparatus as recited in claim 1, wherein said capture system removes meaningless text characters from closed-captioned text of the broadcast, determines a broadcast slot for the broadcast and truncates text lines longer than a predetermined length.

15. An apparatus for analyzing television broadcasts, comprising:

a capture system having a capture machine for each television station and capturing designated television broadcasts, each capture machine removing meaningless text characters from closed-captioned text of the broadcasts, determining broadcast slots for the broadcasts and truncating text lines longer than a predetermined length, each capture machine comprising:

a closed-caption unit receiving capturing the closed captioned text for the broadcast; and

a video/audio capture unit converting a video of the broadcasts into digital images and an audio of the broadcast into digital sound;

an analysis system separating the captured broadcasts into individual stories, analyzing the individual stories and determining competitive characteristics of the broadcast by comparing the closed-captioned text of the stories with statistical information accumulated about prior broadcasts including scoring all segments of text, determining whether the segments can be combined and rescoring the

combined segments, said analysis system scoring and rescoring by comparing the text to a vocabulary table, determining a score table from the vocabulary table and accumulating scores for the text from the score table with the highest score for each classification category becoming a classification for the text, said competitive characteristics comprising one of broadcast topic, talent and production; and

a user interface simultaneously graphically displaying the competitive characteristics of the broadcast and the broadcast for a user.

16. An apparatus. comprising:

means for capturing a designated television broadcast, analyzing the broadcast and determining competitive characteristics of the broadcast;

storage means for storing the broadcast and the characteristics; and

display means for displaying the stored broadcast and the characteristics.

17. A method of analyzing television broadcasts, comprising:

capturing a designated television broadcast; and

analyzing the broadcast and determining competitive characteristics of the broadcast.

18. A storage media storing a process capturing a designated television broadcast and analyzing the broadcast and determining competitive characteristics of the broadcast.

19. A method of television broadcast classification, comprising:

separating the broadcast into stories;

determining the competitive characteristics of a content of the stories; and

creating a statistical database indicating competitive characteristics of story contents.

20. A method of television broadcast classification, comprising:

capturing broadcasts and creating training data; and

creating a statistical database of competitive characteristics of the broadcasts using the training data.

21. A method as recited in claim 20, wherein said statistical database is provided as a tree with branch divisions responsive to at least one of station and time.

22. An apparatus, comprising:

storage storing a television broadcast and competitive characteristics of the broadcast; and

a graphical user interface system graphically displaying the competitive characteristics.

23. An apparatus as recited in claim 22, wherein said system displays the broadcast with the characteristics.

24. An apparatus as recited in claim 22, wherein said storage stores ratings data for a broadcast and said system displays the characteristics and the ratings data.

25. An apparatus as recited in claim 22, wherein said storage stores plural broadcasts and corresponding competitive characteristics of the broadcasts and said system simultaneously displays the corresponding characteristics.

26. An apparatus as recited in claim 25, wherein said system simultaneously plays the plural broadcasts.

27. An apparatus as recited in claim 26, wherein the system synchronizes the broadcasts to a one of the broadcasts whose audio is being played.

28. An apparatus as recited in claim 22, further comprising a capture unit providing the broadcast and the competitive characteristics.

29. An apparatus as recited in claim 28, wherein said system provides an ability to select at least one of television station, date of capture and time of capture.

30. An apparatus as recited in claim 28, wherein said system provides an ability to display at least one of broadcast pacing, story length, ratio of news broadcast time to advertisement broadcast time, a list of story segments indicating competitive characteristics of the segments, time related viewer retention data and lead-in related viewer retention data.

31. An apparatus as recited in claim 23, wherein said system performs non-linear playback of video segments.

32. An apparatus comprising:

a capture system capturing a designated television broadcast, analyzing the broadcast and determining competitive characteristics of the broadcast, said capture system having a capture machine for each television station and capturing designated television broadcasts, each capture machine removing meaningless text characters from closed-captioned text of the broadcasts, determining broadcast slots for the broadcasts and truncating text lines longer than a predetermined length, separating the captured broadcasts into individual stories, analyzing the individual stories and determining competitive characteristics of the broadcast by comparing the closed-captioned text of the stories with statistical information accumulated about prior broadcasts including scoring all segments of text, determining whether the segments can be combined and rescoring the combined segments, said analysis system scoring and rescoring by comparing the text to a vocabulary table, determining a score table from the vocabulary table and accumulating scores for the text from the score table with the highest score for each classification category becoming a classification for the text, and each capture machine comprising: a closed-caption unit receiving capturing the closed captioned text for the broadcast; and a video/audio capture unit converting a video of the broadcasts into digital images and an audio of the broadcast into digital sound; a storage system storing the television broadcasts, competitive characteristics of the broadcasts and ratings data for the broadcasts; a graphical user interface system graphically displaying the competitive characteristics comprising topic, talent, production, broadcast pacing, story length, ratio of news broadcast time to advertisement broadcast time, time related viewer retention data and lead-in related viewer retention data, playing the stored broadcasts and displaying the ratings data for the broadcasts; and a storage media storing a process of the capture machine performing the capturing a designated television broadcast and the analyzing the broadcast and the determining competitive characteristics of the broadcast.

33. A graphical user interface, comprising:

a video display region displaying a video broadcast; and

a data display region, simultaneously displayed with said video display region and in association therewith, displaying one or more competitive characteristics of the video broadcast.

34. An interface as recited in claim 33, wherein said characteristics comprise one of topic, talent, production and ratings.

35. An interface as recited in claim 33, wherein said characteristics are displayed as a chart.

36. An interface as recited in claim 33, wherein said interface further comprises a selection display region, simultaneously displayed with said video display region and in association therewith, displaying one or more display characteristics from among station, date and time.

37. A broadcast video competitive analysis audio/video graphical user interface, comprising:

a video display region capable of simultaneously displaying two or more video broadcasts;

an audio signal;

a data display region, simultaneously displayed with said video display region and in association therewith, displaying a chart of competitive characteristics of the video broadcast of the video broadcasts; and

a selection display region, simultaneously displayed with said video display region and in association therewith, displaying one or more display characteristics from among station, date and time.

38. A computer readable storage medium comprising a vocabulary table storing phrases of predetermined lengths, a phrase identifier and a score table identifier and a phrase score table including entries for each phrase identifier, an occurrence frequency of the phrase in different phrase categories.

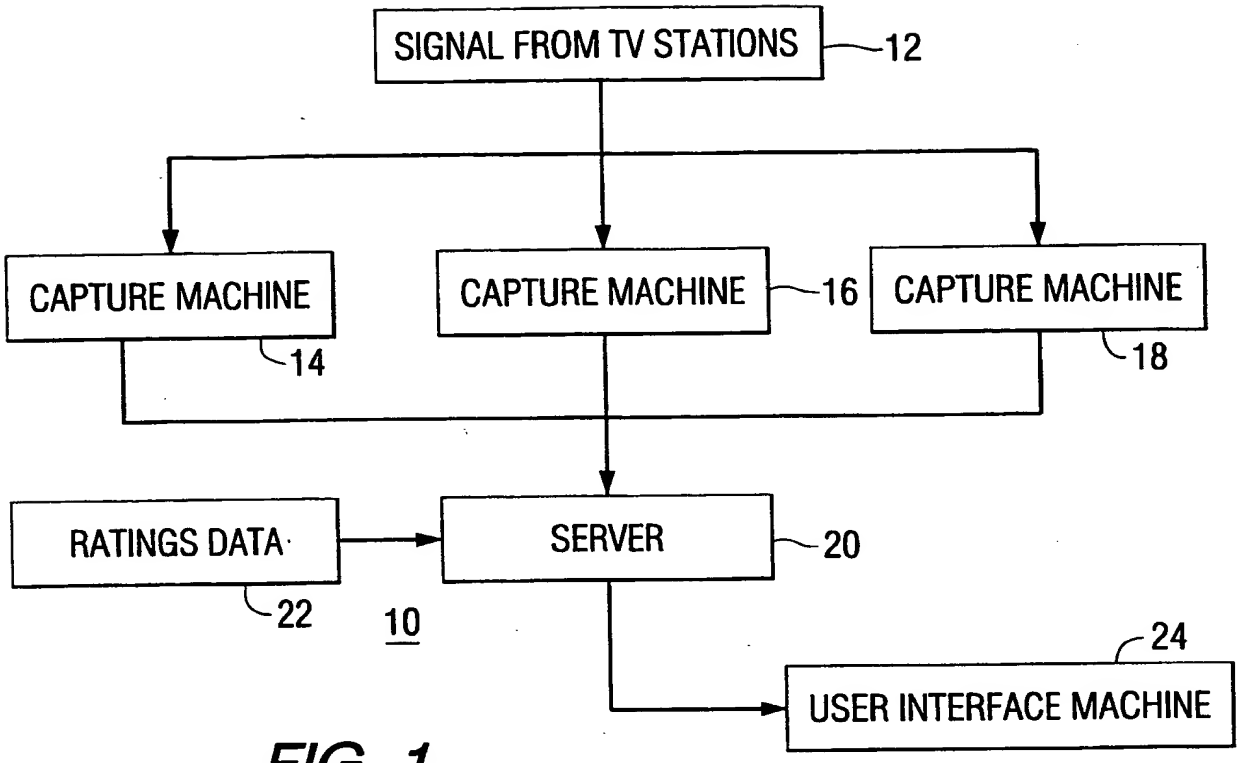

SIGNAL FROM TV STATIONS
12
CAPTURE MACHINE
14
CAPTURE MACHINE
16
CAPTURE MACHINE
18
RATINGS DATA
22
SERVER
20
10
24
USER INTERFACE MACHINE

FIG. 1

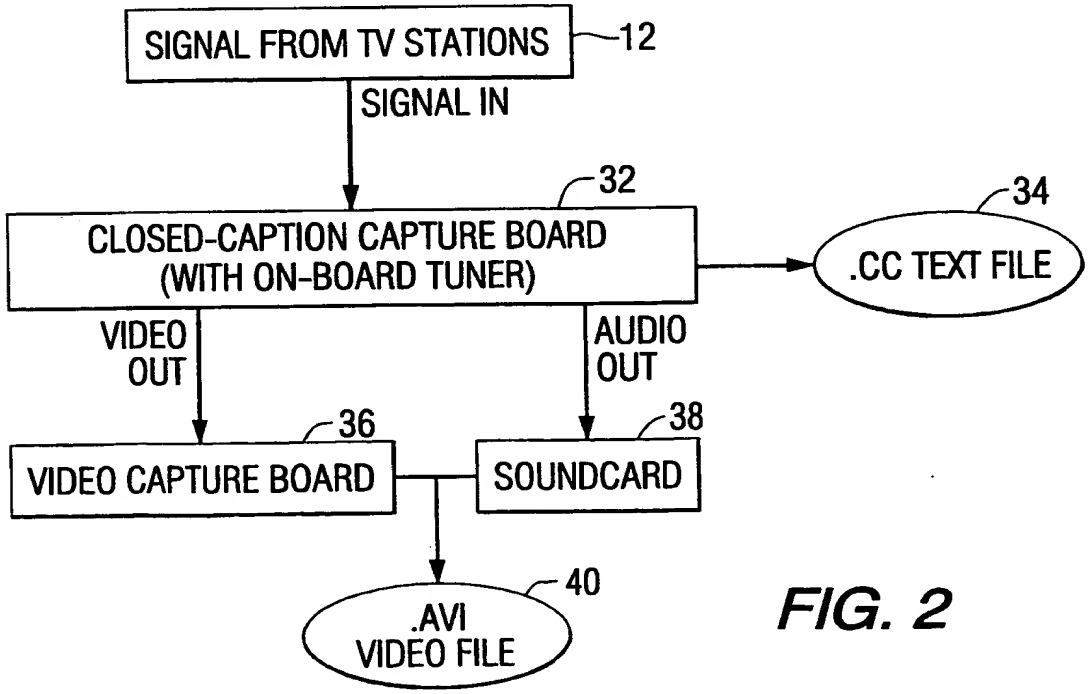

SIGNAL FROM TV STATIONS
12
SIGNAL IN
32
CLOSED-CAPTION CAPTURE BOARD
(WITH ON-BOARD TUNER)
34
.CC TEXT FILE
VIDEO
OUT
AUDIO
OUT
36
VIDEO CAPTURE BOARD
38
SOUNDCARD
40
.AVI
VIDEO FILE

FIG. 2

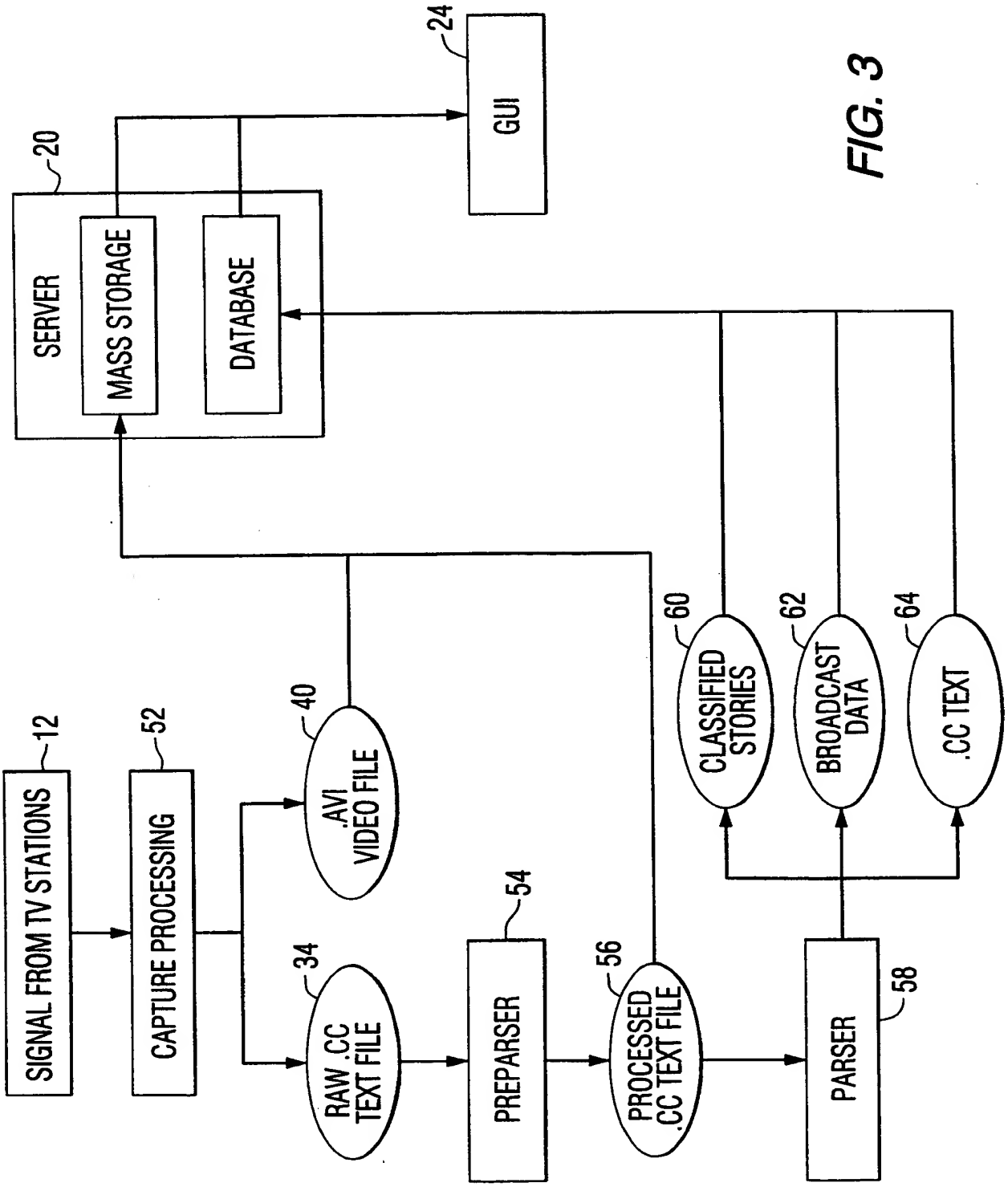

SIGNAL FROM TV STATIONS
12
CAPTURE PROCESSING
52
.AVI VIDEO FILE
40
RAW .CC TEXT FILE
34
PREPARSER
54
PROCESSED .CC TEXT FILE
56
PARSER
58
CLASSIFIED STORIES
60
BROADCAST DATA
62
.CC TEXT
64
SERVER
MASS STORAGE
DATABASE
20
GUI
24

FIG. 3

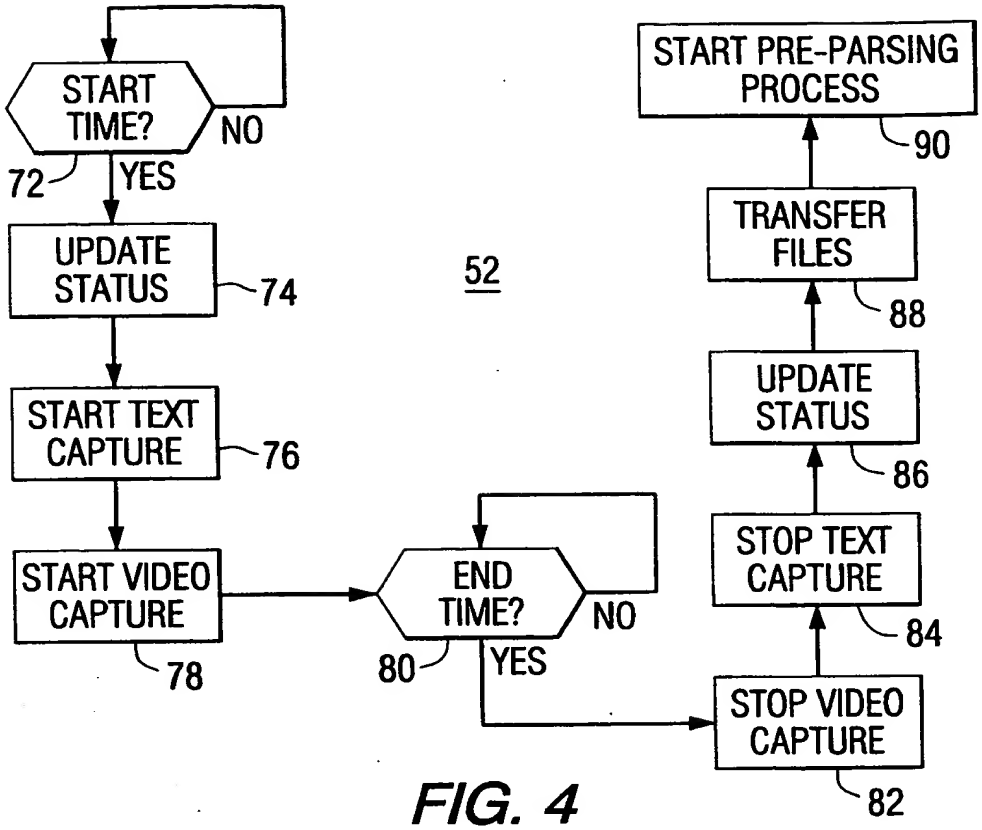

START TIME?
NO
YES
72
UPDATE STATUS
74
START TEXT CAPTURE
76
START VIDEO CAPTURE
78
END TIME?
NO
YES
80
STOP VIDEO CAPTURE
82
STOP TEXT CAPTURE
84
UPDATE STATUS
86
TRANSFER FILES
88
START PRE-PARSING PROCESS
90
52

FIG. 4

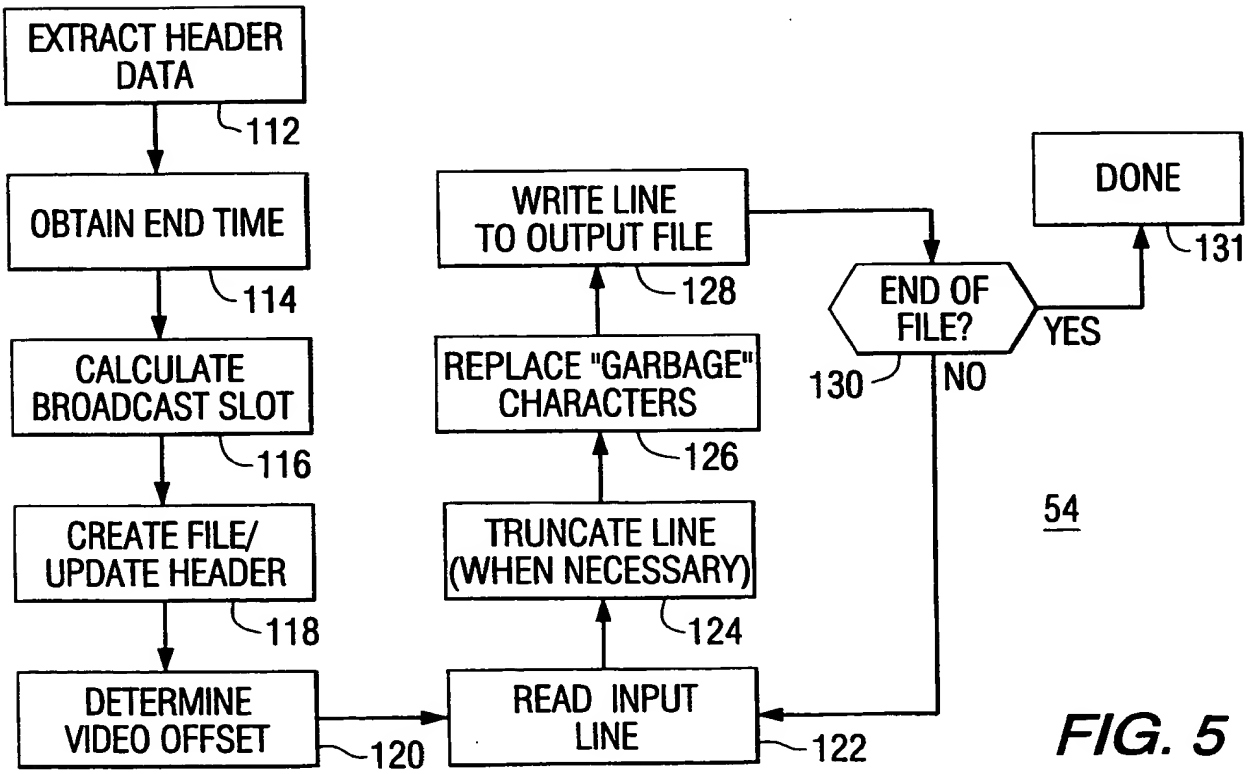

EXTRACT HEADER DATA
112
OBTAIN END TIME
114
CALCULATE BROADCAST SLOT
116
CREATE FILE/ UPDATE HEADER
118
DETERMINE VIDEO OFFSET
120
READ INPUT LINE
122
TRUNCATE LINE (WHEN NECESSARY)
124
REPLACE "GARBAGE" CHARACTERS
126
WRITE LINE TO OUTPUT FILE
128
END OF FILE?
YES
NO
130
DONE
131
54

FIG. 5

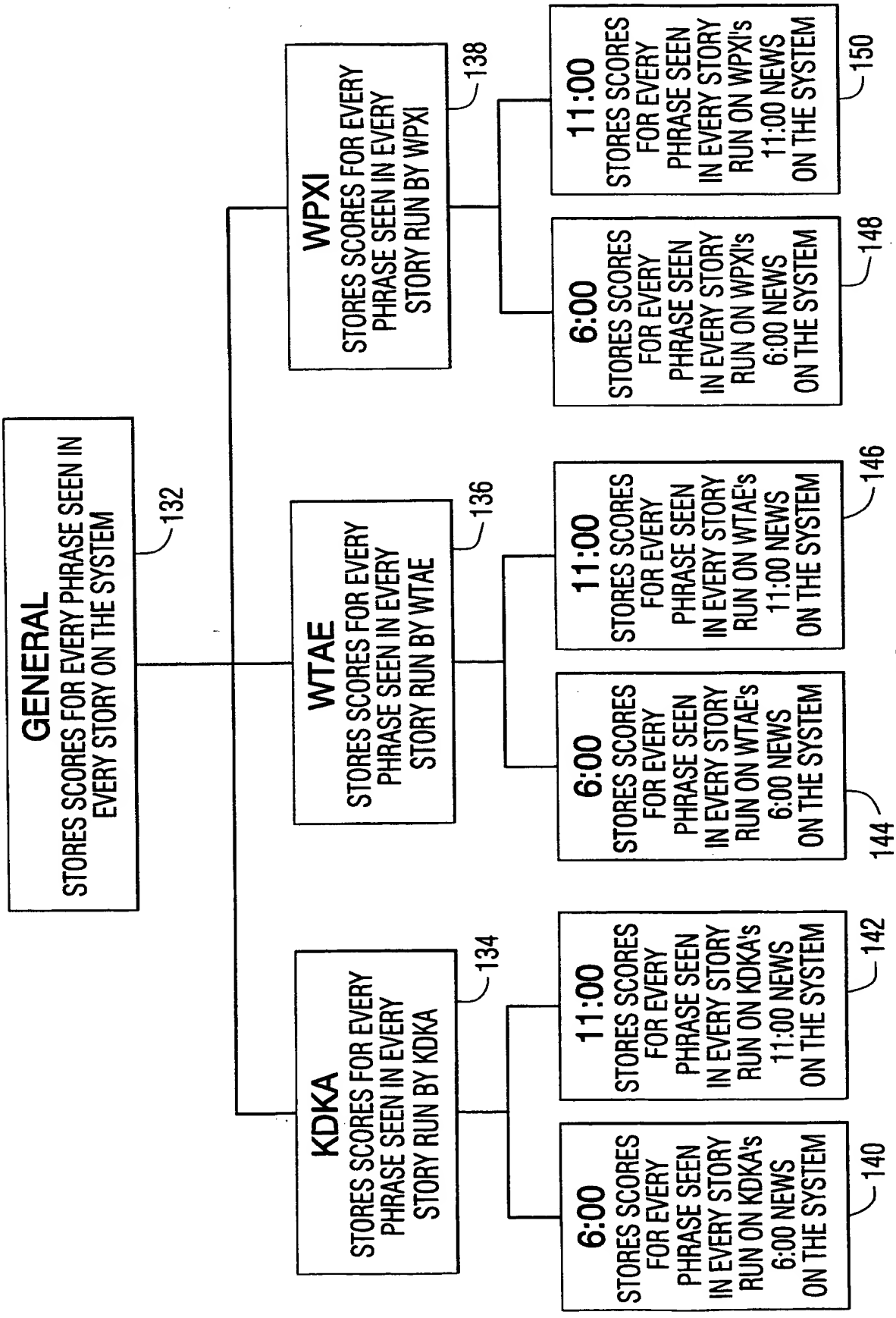

GENERAL
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY ON THE SYSTEM
132
KDKA
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY RUN BY KDKA
134
WTAE
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY RUN BY WTAE
136
WPXI
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY RUN BY WPXI
138
6:00
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY RUN ON KDKA's 6:00 NEWS ON THE SYSTEM
140
11:00
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY RUN ON KDKA's 11:00 NEWS ON THE SYSTEM
142
6:00
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY RUN ON WTAE's 6:00 NEWS ON THE SYSTEM
144
11:00
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY RUN ON WTAE's 11:00 NEWS ON THE SYSTEM
146
6:00
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY RUN ON WPXI's 6:00 NEWS ON THE SYSTEM
148
11:00
STORES SCORES FOR EVERY PHRASE SEEN IN EVERY STORY RUN ON WPXI's 11:00 NEWS ON THE SYSTEM
150

FIG. 6

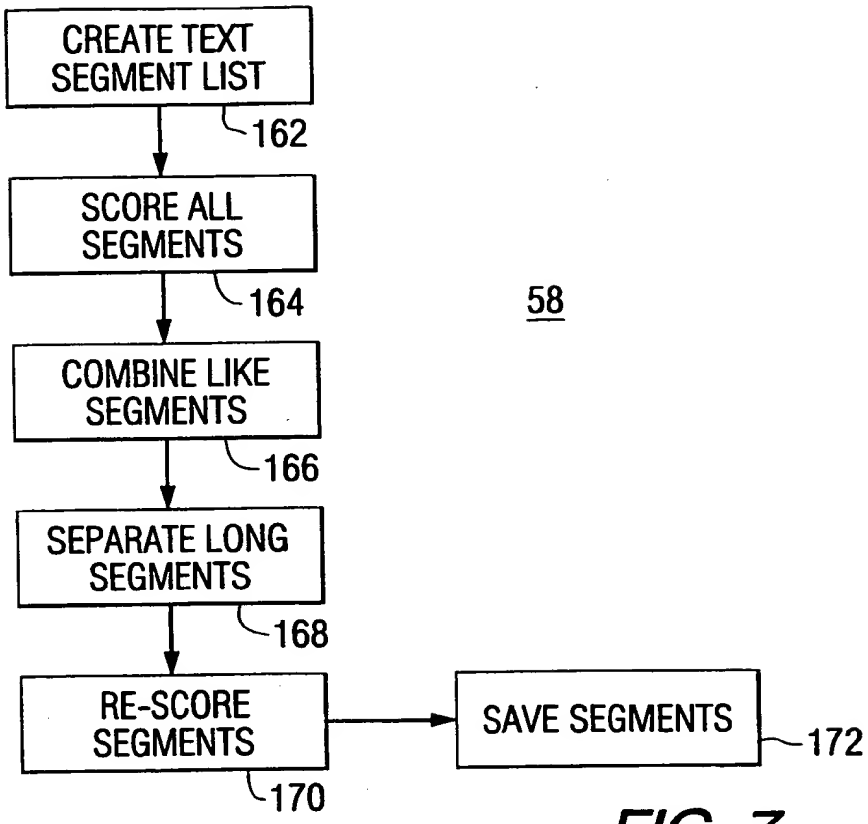

CREATE TEXT SEGMENT LIST
162
SCORE ALL SEGMENTS
164
58
COMBINE LIKE SEGMENTS
166
SEPARATE LONG SEGMENTS
168
RE-SCORE SEGMENTS
170
SAVE SEGMENTS
172

FIG. 7

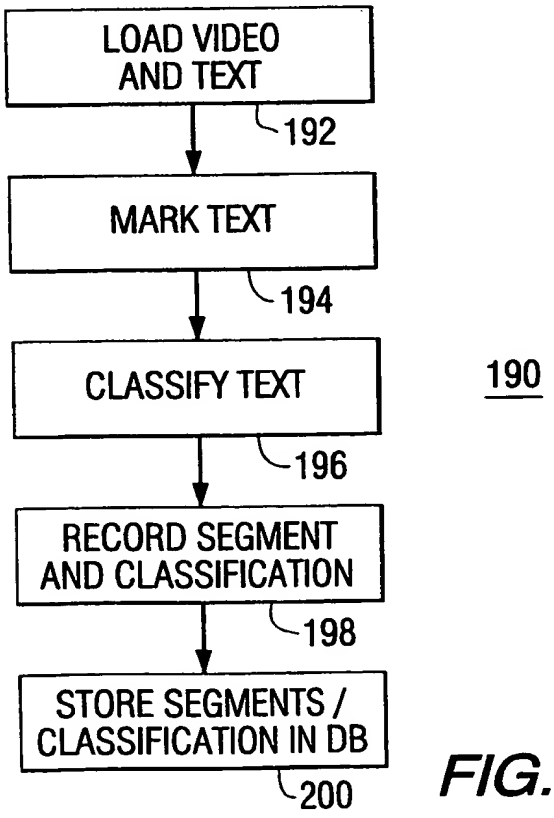

LOAD VIDEO AND TEXT
192
MARK TEXT
194
CLASSIFY TEXT
190
196
RECORD SEGMENT AND CLASSIFICATION
198
STORE SEGMENTS / CLASSIFICATION IN DB
200

FIG. 8

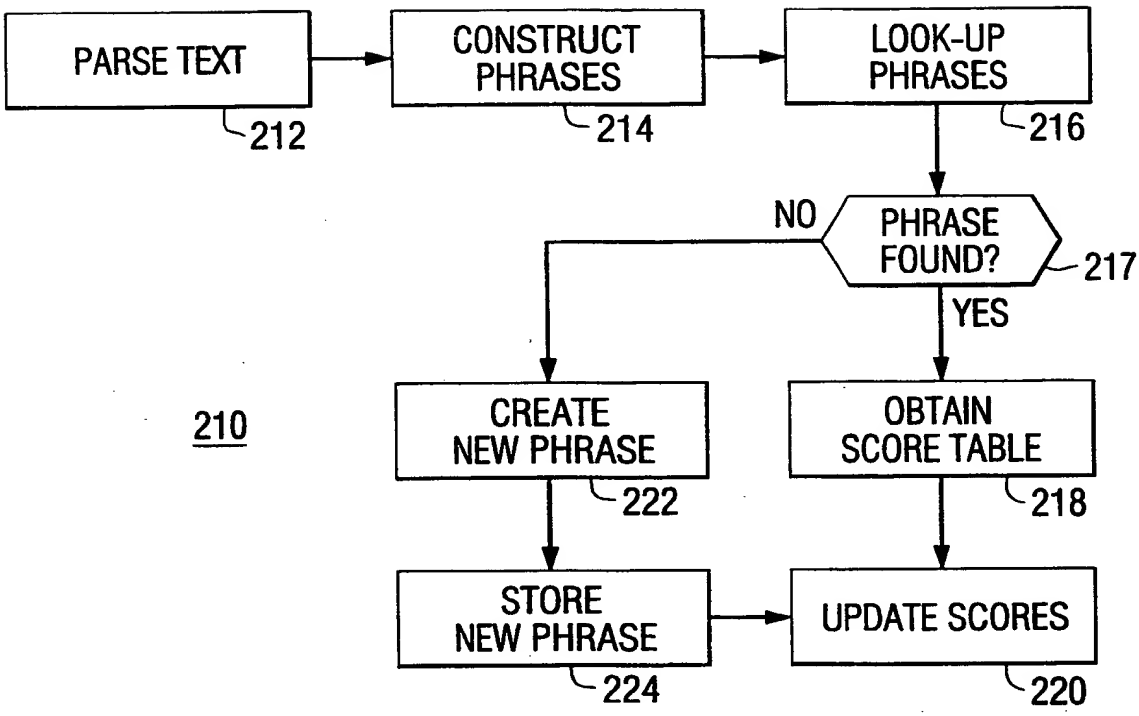

PARSE TEXT
212
CONSTRUCT PHRASES
214
LOOK-UP PHRASES
216
NO
PHRASE FOUND?
217
YES
210
CREATE NEW PHRASE
222
OBTAIN SCORE TABLE
218
STORE NEW PHRASE
224
UPDATE SCORES
220

FIG. 9

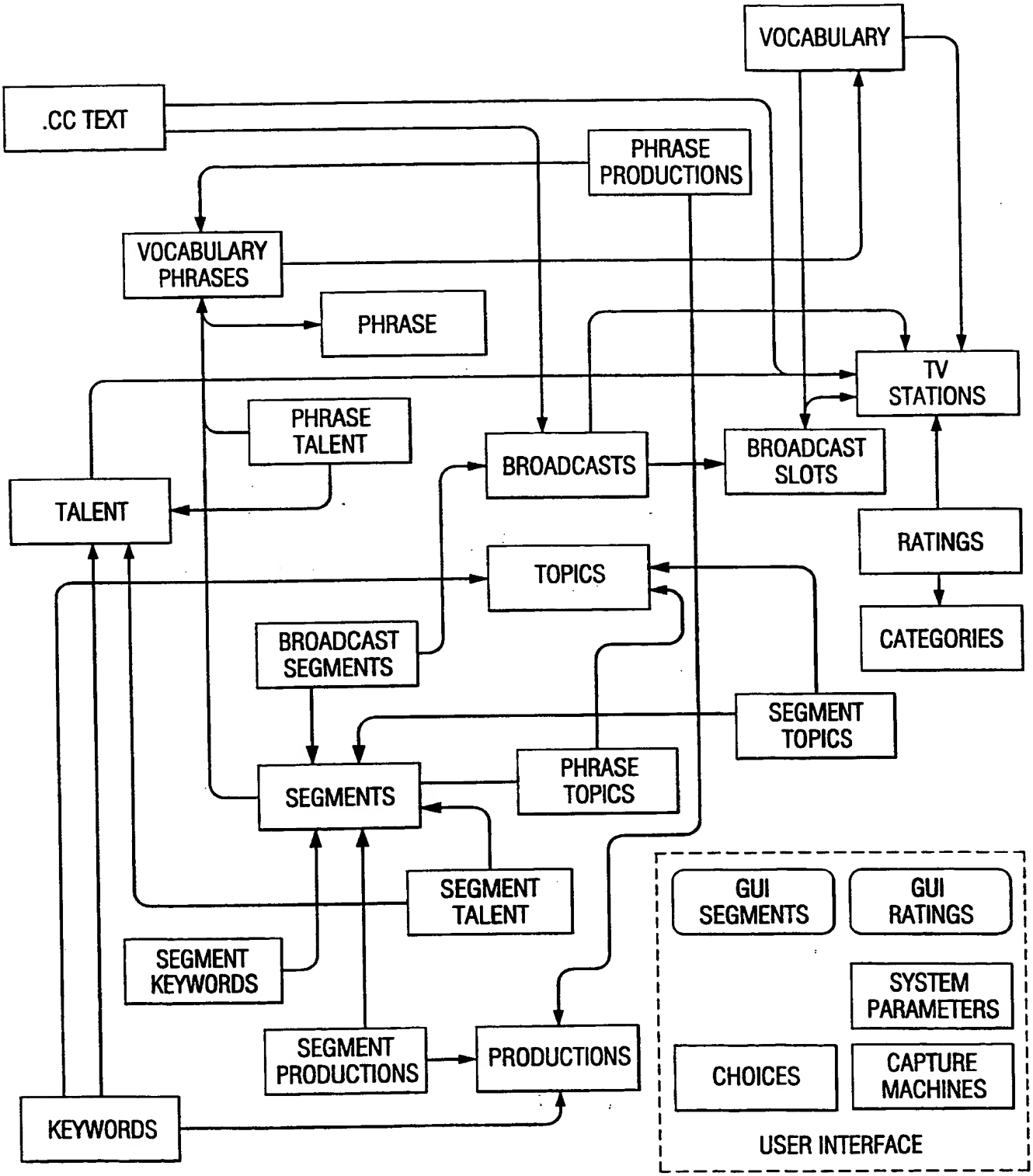

VOCABULARY
.CC TEXT
PHRASE PRODUCTIONS
VOCABULARY PHRASES
PHRASE
TV STATIONS
PHRASE TALENT
BROADCASTS
BROADCAST SLOTS
TALENT
RATINGS
TOPICS
CATEGORIES
BROADCAST SEGMENTS
SEGMENT TOPICS
SEGMENTS
PHRASE TOPICS
SEGMENT TALENT
GUI SEGMENTS
GUI RATINGS
SEGMENT KEYWORDS
SYSTEM PARAMETERS
SEGMENT PRODUCTIONS
PRODUCTIONS
CHOICES
CAPTURE MACHINES
KEYWORDS
USER INTERFACE

FIG. 10

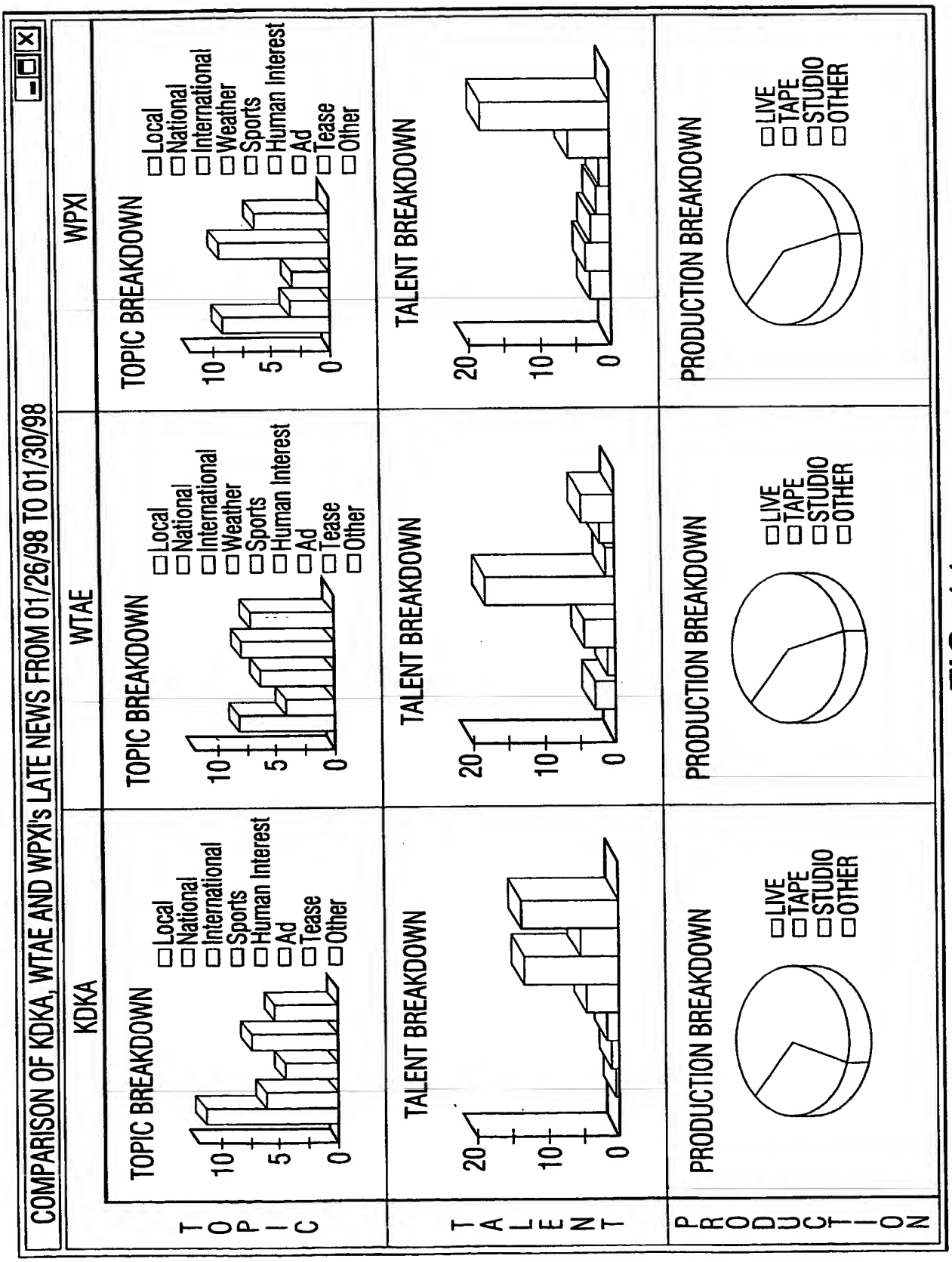

COMPARISON OF KDKA, WTAE AND WPXI's LATE NEWS FROM 01/26/98 TO 01/30/98
KDKA
WTAE
WPXI
TOPIC
TALENT
PRODUCTION
TOPIC BREAKDOWN
Local
National
International
Sports
Human Interest
Ad
Tease
Other
Weather
TALENT BREAKDOWN
PRODUCTION BREAKDOWN
LIVE
TAPE
STUDIO
OTHER

FIG. 11

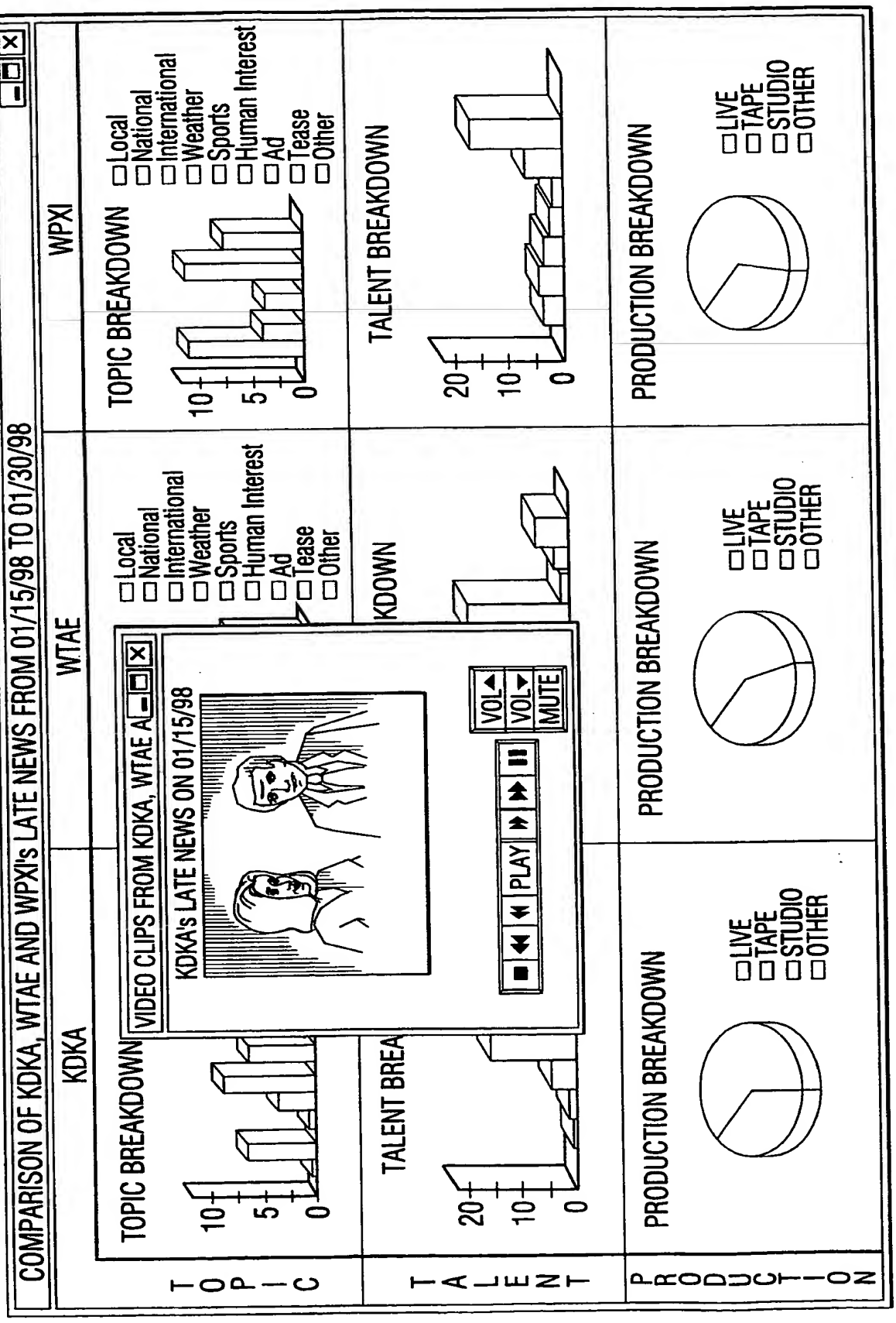

COMPARISON OF KDKA, WTAE AND WPXI's LATE NEWS FROM 01/15/98 TO 01/30/98
KDKA
WTAE
WPXI
TOPIC
TALENT
PRODUCTION
TOPIC BREAKDOWN
Local
National
International
Weather
Sports
Human Interest
Ad
Tease
Other
VIDEO CLIPS FROM KDKA, WTAE A
KDKA's LATE NEWS ON 01/15/98
PLAY
VOL▲
VOL▼
MUTE
TALENT BREAKDOWN
KDOWN
PRODUCTION BREAKDOWN
LIVE
TAPE
STUDIO
OTHER

FIG. 12

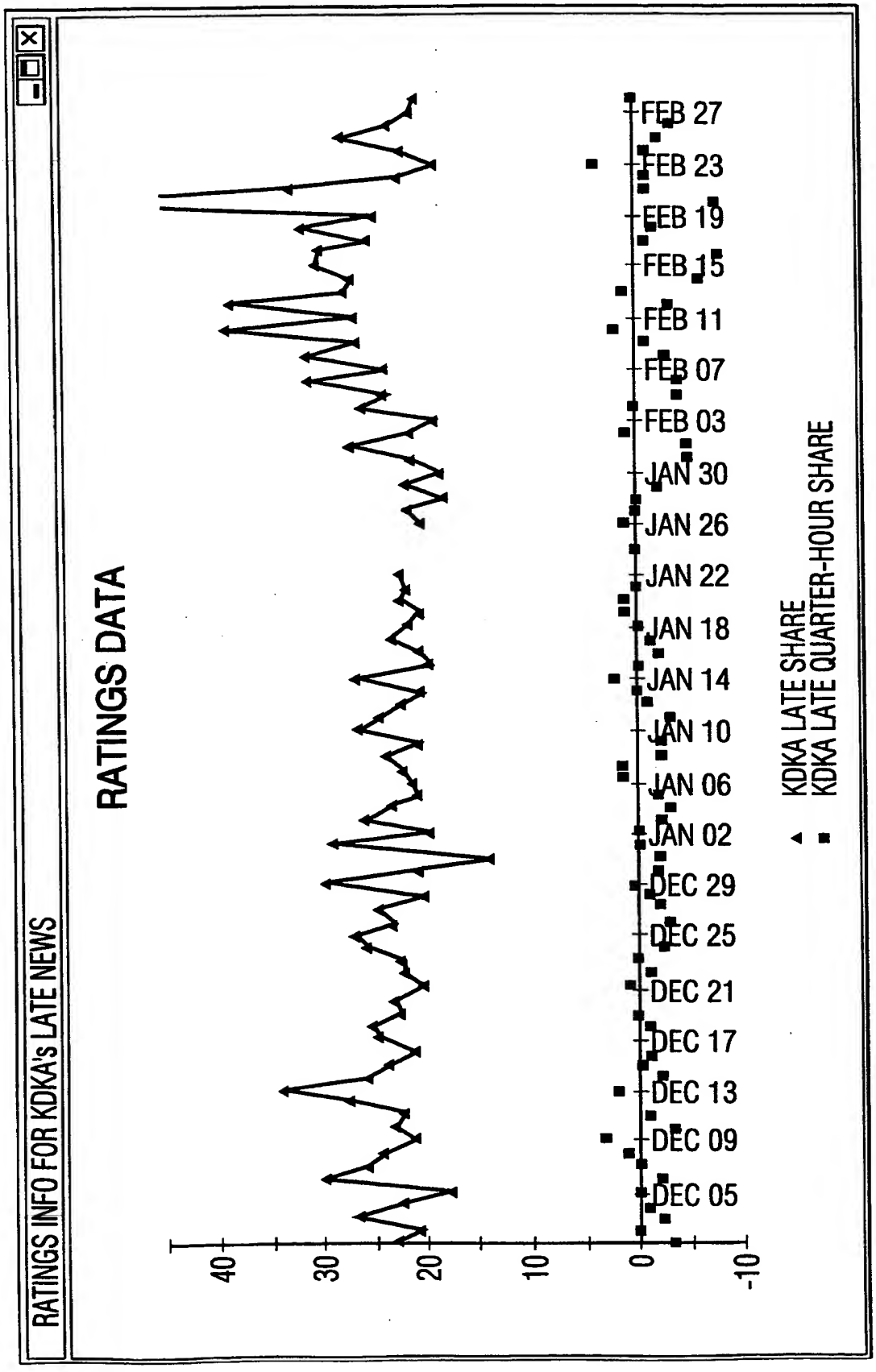

RATINGS INFO FOR KDKA's LATE NEWS
RATINGS DATA
FEB 27
FEB 23
FEB 19
FEB 15
FEB 11
FEB 07
FEB 03
JAN 30
JAN 26
JAN 22
JAN 18
JAN 14
JAN 10
JAN 06
JAN 02
DEC 29
DEC 25
DEC 21
DEC 17
DEC 13
DEC 09
DEC 05
40
30
20
10
0
-10
KDKA LATE SHARE
KDKA LATE QUARTER-HOUR SHARE

FIG. 13

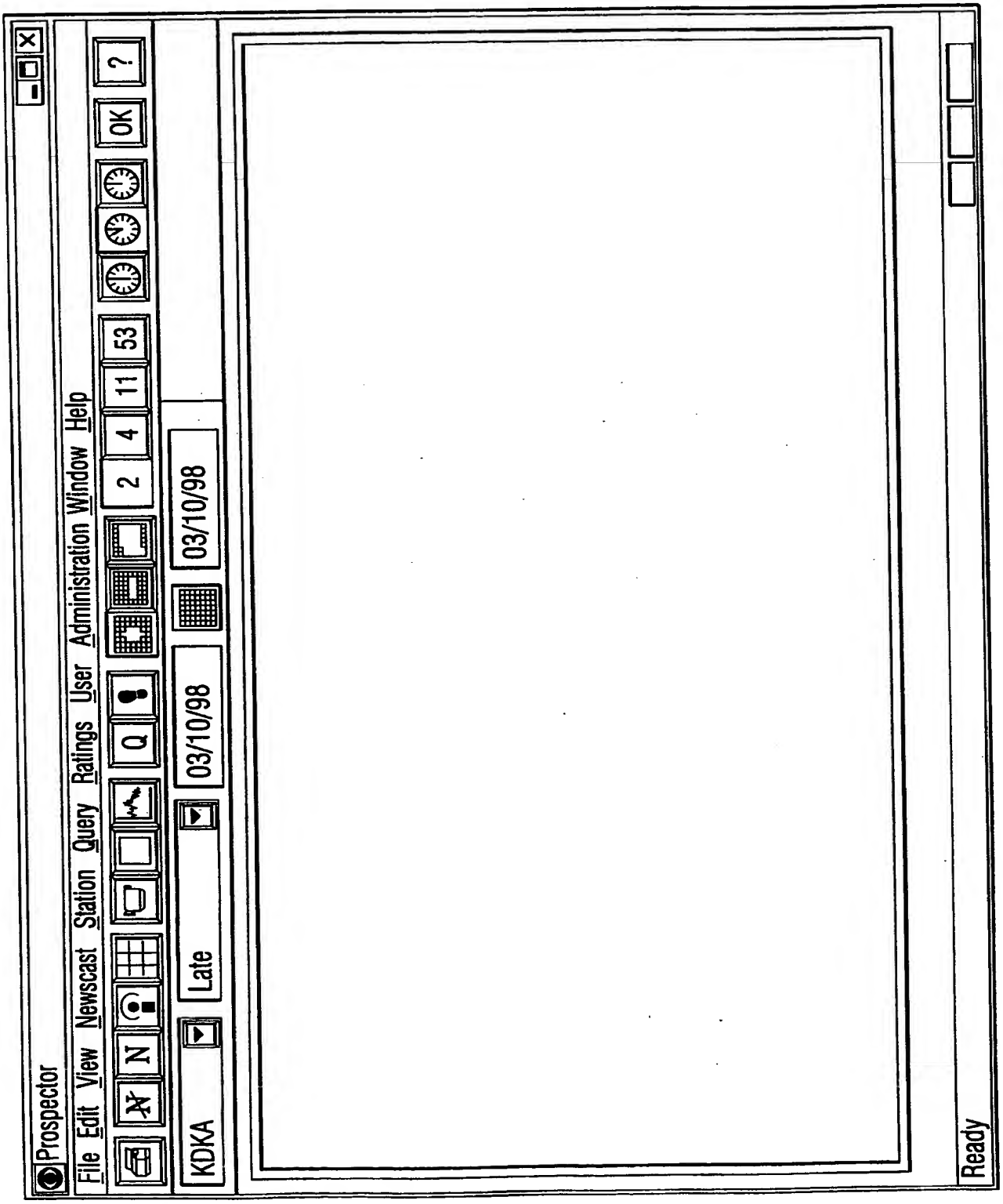

Prospector
File Edit View Newscast Station Query Ratings User Administration Window Help
KDKA
Late
03/10/98
03/10/98
OK
Ready

FIG. 14

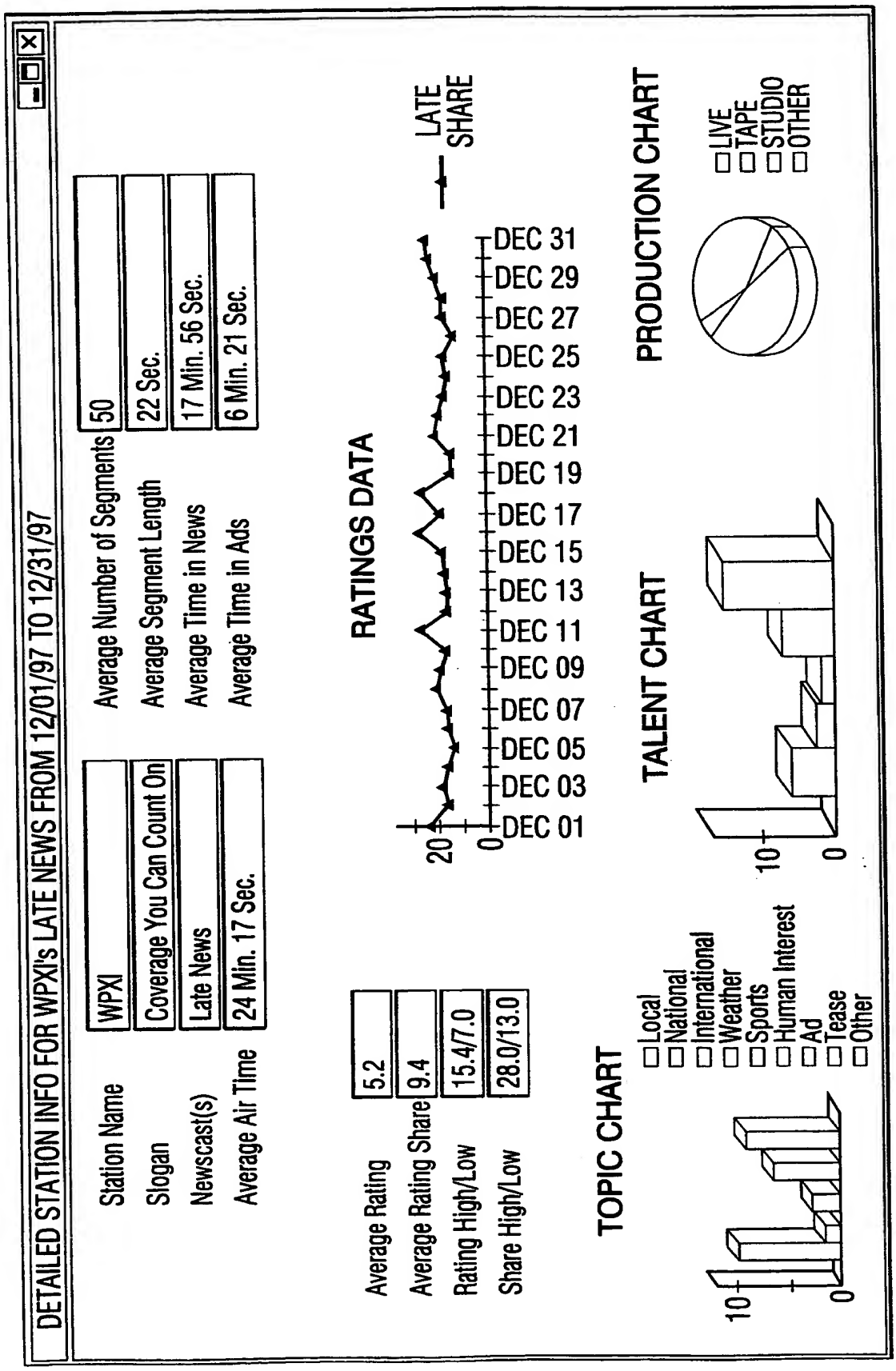

DETAILED STATION INFO FOR WPXI's LATE NEWS FROM 12/01/97 TO 12/31/97
Station Name
WPXI
Slogan
Coverage You Can Count On
Newscast(s)
Late News
Average Air Time
24 Min. 17 Sec.
Average Number of Segments
50
Average Segment Length
22 Sec.
Average Time in News
17 Min. 56 Sec.
Average Time in Ads
6 Min. 21 Sec.
Average Rating
5.2
Average Rating Share
9.4
Rating High/Low
15.4/7.0
Share High/Low
28.0/13.0
RATINGS DATA
LATE SHARE
DEC 31
DEC 29
DEC 27
DEC 25
DEC 23
DEC 21
DEC 19
DEC 17
DEC 15
DEC 13
DEC 11
DEC 09
DEC 07
DEC 05
DEC 03
DEC 01
20
0
TOPIC CHART
Local
National
International
Weather
Sports
Human Interest
Ad
Tease
Other
10
0
TALENT CHART
10
0
PRODUCTION CHART
LIVE
TAPE
STUDIO
OTHER

FIG. 15

INTERNATIONAL SEARCH REPORT

In' national Application No
PCT/US 99/16799

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04H9/00 H04N7/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04H H04N G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate. of the relevant passages	Relevant to claim No.
X	US 5 371 673 A (FAN DAVID P) 6 December 1994 (1994-12-06) abstract column 25, line 25 - line 66 claims 1,4,6 figure 2 ---	1-38
X	GB 2 258 065 A (MCCONNELLS DEVELOPMENTS LIMITE) 27 January 1993 (1993-01-27) the whole document ---	1-38
X	US 4 550 407 A (COUASNON TRISTAN DE ET AL) 29 October 1985 (1985-10-29) abstract claims 1,4,6 figures 3,4 --- -/--	1,16-18, 22,23

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents. such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 November 1999	18/11/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Simon, V

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/16799

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate. of the relevant passages	Relevant to claim No.
A	US 4 857 999 A (WELSH RUSSELL J) 15 August 1989 (1989-08-15) abstract claim 1 figures 1,4,5	1,16-18, 22,23,38

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/16799

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5371673 A	06-12-1994	US 4930077 A	29-05-1990
GB 2258065 A	27-01-1993	BE 1003154 A	10-12-1991
US 4550407 A	29-10-1985	FR 2529041 A	23-12-1983
		CA 1208759 A	29-07-1986
		EP 0099276 A	25-01-1984
		JP 59016479 A	27-01-1984
US 4857999 A	15-08-1989	CA 1332462 A	11-10-1994
		CA 1337363 A	17-10-1995

This Page Blank (uspto)